VTONDemand: A Framework for Indexing, Searching and On-Demand Playback of RTP-Based Multimedia Conferences

Baldine B. Paul, Dave Gibbon, Glenn Cash and M. Reha Civanlar
AT&T Labs - Research
Red Bank NJ 07701-7033
Phone: 732-345-3312 Fax: 732-345-3033
{baldine,deg,glenn,mrc}@research.att.com

Abstract - In this paper, we describe the implementation of a system for reliable recording and on-demand indexed playback of multimedia conferences using the services provided by the RTP/RTCP protocols over an Intranet. Implementation issues will be underlined regarding the file structure to support random access during playback, indexing techniques and database searching capabilities.

INTRODUCTION

In this paper, we describe the implementation of a system for reliable recording and on-demand indexed playback of multimedia conferences using the services provided by the RTP/RTCP protocols over an Intranet. Implementation issues will be underlined regarding the file structure to support random access during playback, indexing techniques and database searching capabilities.

RELATED WORK

We described in [1] the AT&T VideoTalks™ Jukebox System that is used within an Intranet for the recording, databasing and indexing of technical presentations and other meetings of general interest for later on-demand playback on the desktop systems (Unix, Windows95/NT). Similar tools have been developed in recent years to provide recording and playback capabilities for video conferencing sessions transmitted through the Internet’s Multicast Backbone known as MBone. Examples of these are rtmp [3], the MBone VCR [5], the Interactive Multimedia Jukebox System, IMJ [6] and the Multicast Media-on-Demand System, mMOD [7].
THE VTONDEMAND FRAMEWORK

VTonDemand is a framework for utilizing the VTJukebox multimedia database with random access, media browsing, navigating, and searching capabilities. As shown in Figure 1, the VideoTalksTM Jukebox system, VTJukebox, comprises: (i) an HTTP interface on the user's desktop for requesting playback of recorded sessions, (ii) a centrally located server for database management and session playback/record scheduling. (iii) one or more RTP-recorder/player controlled by the centrally located server; the actual recording or playback of multimedia sessions is done by this recorder/player. (iv) a client or set of client applications to receive and display the different media on a user's desktop. For audio and video playback, there are commercially available players like Cisco IPTV and Icast, or the freely available MBone tools vic and vat. For viewgraph and pointer reception and display, we have developed our own JPEG-based viewgraph decoder [10].

VTonDemand Navigation

VTonDemand supports both linear and non-linear playback of the recorded sessions. Specifically, it supports VCR-type capabilities such as fast-forward and rewind and also playback of a subset of the recorded media with full user control. For example, the user can start, stop, pause the playback session, select to playback audio-only, or audio and slides only and navigate through the selected media.

To facilitate control of the playback session, a rendering engine generates application-specific HTML user interface views of the selected talk using pre-defined templates [9]. The user is presented with icon images of viewgraphs serving as anchor points into the videoconferencing session, in addition to the standard VCR-like control buttons. By clicking on any of these viewgraph captions, displayed on a web page, the user can start or advance the play position to a point of interest in the presentation.

Selection of an image caption by the user translates into a playback query having as argument the offset of this particular viewgraph in the recorded multimedia file. The VTonDemand server maintains a cross indexing file associating viewgraph offset with session playback time. The viewgraphs to be used in the index page are generated from the recorded high resolution JPEG-encoded viewgraphs.

Session Description File

Information about each recorded session is maintained at the central scheduler in a session description file. The session description follows the syntax of the Session Description Protocol, SDP [4], with some proprietary extensions. For instance, information recorded in the SDP file include the name of the presenter, the date, time and duration of the presentation and an abstract of
Figure 1: The VTJukebox System.
the presentation, if available. This information is gathered by the scheduler at the time the recording is requested.

The information embedded in the SDP file provides direct support for various searching modalities for VTonDemand. For instance, we can search by date, presenter’s first and last name and keywords within the embedded abstract. Such search queries are handled by searching through the session description files using a text-based search agent. The result of the search is a string containing the name and path of the corresponding session in the VTJukebox database.

**Multimedia File Format**

Linear and non-linear playback capabilities for VTonDemand are supported by the file indexing strategy developed for the recorded multimedia data [1]. This file format allows easy forward and backward traversing, within a particular medium and across multiple media.

Each RTP/RTCP packet received from the network is prepended with a recorder packet header containing the reception time for the packet, the original time stamp and the indexing information. The indexing information includes the size of the packet, the offsets for the following and the preceding packets in the multimedia file and the offsets of the previous and the next RTP-packet for the current medium. The indexing information is used to traverse the media file and, depending on the user’s request, playback audio-only, viewgraph-only or any combination of the available media.

**Viewgraph Icons Generation and Selection**

During the live presentation, image analysis takes place at the viewgraph encoder in order to detect changes in viewgraph content. This analysis assists in deciding whether to retransmit a buffered viewgraph or encode a new one for transmission. In this context, the goal is to reproduce the experience of attending the presentation while minimizing the average bitrate of the viewgraph stream. In this process, viewgraphs that are significantly different from preceding ones are marked as possible key viewgraphs that will serve as seeds to the icons generation process.

The criterion for selecting a viewgraph for icon generation differs from that of viewgraph selection for transmission. The goal of the icon selection process is to provide a compact representation of the visual contents of the presentation. In this context, we want to reduce as much as possible redundancy and repetition of similar icons. The problem at hand is to minimize the number of frames retained while maximizing the visual information content conveyed by the retained frames.

We have implemented a multi-stage approach to the problem of determining which slides to transmit, and which to retain as icons. First, an efficient correlation-based image comparison with noise tolerance is used to determine
if a slide should be transmitted. The next stage operates on the set of slides that have been identified by the first stage. In this second stage, a temporal filter is used to eliminate transient frames that occur when the viewgraph is being positioned, for example. A final stage removes any remaining blank images that may occur between viewgraphs. A first pass through these stages is done online during viewgraph encoding. Then, the emphasis is placed on speed to meet the real-time requirements. A second pass is completed offline on the recorded viewgraphs.

**Results**

While this automatic indexing cannot compete with the quality of human judgment for icon selection, it produces an acceptable set of icon images for visually browsing a VideoTalk. Figure 2 shows the temporal characteristics of a particular technical talk that we examined. In this example, 501 images were detected and transmitted as described in the first stage of processing. The plot shows the number of images retained from the first stage of processing for periods of time given on the X-axis. For example, if we filter out all viewgraphs that have appeared for less than five seconds, we move from 501 images to only 120. In fact, if a particular application requires a fixed maximum number of icon images, then this filtering method will produce the set that maximizes the time-coverage. Based on our experience, we find that filtering out viewgraphs that span 8 seconds or less produces good results. For the example shown here, the final representation included 58 images after temporal filtering, removing blank images, and some manual editing.

![Figure 2: Number of icons generated as a function of the allowable minimum presentation period.](image-url)
CONCLUSIONS

In this paper, we presented a prototype system for the recording and on-demand playback of RTP-based multimedia conferences with indexing and searching capabilities. Additional work is ongoing to improve the user interface, introduce additional functionalities and improve the efficiency of the automatic indexing of the video-conferences. We are also developing HTML-based tools to allow authors to further edit the results of the automatic indexing.

References


